

**LAS HUELLAS DEL FUTURO ESTÁN
EN LA WEB: CONSTRUCCIÓN
DE INDICADORES ADELANTADOS
A PARTIR DE LAS BÚSQUEDAS
EN INTERNET**

2012

Concha Artola y Enrique Galán

**Documentos Ocasionales
N.º 1203**

BANCO DE ESPAÑA
Eurosistema



**LAS HUELLAS DEL FUTURO ESTÁN EN LA WEB: CONSTRUCCIÓN DE
INDICADORES ADELANTADOS A PARTIR DE LAS BÚSQUEDAS EN INTERNET**

LAS HUELLAS DEL FUTURO ESTÁN EN LA WEB: CONSTRUCCIÓN DE INDICADORES ADELANTADOS A PARTIR DE LAS BÚSQUEDAS EN INTERNET (*)

Concha Artola y Enrique Galán (**)

BANCO DE ESPAÑA

(*) Nos gustaría agradecer los comentarios de Juan Peñalosa, Eloísa Ortega y Patry Tello; el apoyo técnico de Coral García Esteban, Roberto López Pavón y Domingo Pérez Cañete; la labor de edición de María Jesús Vázquez; y, muy especialmente, la ayuda de Agustín Maravall.

(**) Direcciones de contacto: artola@bde.es; enrique.galan@bde.es.

La serie de Documentos Ocasionales tiene como objetivo la difusión de trabajos realizados en el Banco de España, en el ámbito de sus competencias, que se consideran de interés general.

Las opiniones y análisis que aparecen en la serie de Documentos Ocasionales son responsabilidad de los autores y, por tanto, no necesariamente coinciden con los del Banco de España o los del Eurosistema.

El Banco de España difunde sus informes más importantes y la mayoría de sus publicaciones a través de la red Internet en la dirección <http://www.bde.es>.

Se permite la reproducción para fines docentes o sin ánimo de lucro, siempre que se cite la fuente.

© BANCO DE ESPAÑA, Madrid, 2012

ISSN: 1696-2230 (edición electrónica)

Resumen

En este trabajo se revisan algunas de las aplicaciones que utilizan la vasta información proporcionada por las búsquedas realizadas por los usuarios de Internet para el análisis y la previsión económica. Este enorme volumen de información, disponible en tiempo real, es manejable para los analistas gracias a herramientas estadísticas como “Google Insights for Search” que permiten clasificar y evaluar las tendencias en distintas áreas de interés. Los trabajos precedentes se han centrado predominantemente en el mercado de trabajo, en los mercados de la vivienda, en las ventas minoristas y en la confianza de los consumidores. En este trabajo se ha presentado una aplicación muy específica para la economía española: los flujos de turistas británicos entrados en España —los principales clientes de la industria turística española—. Los modelos de corto plazo que incluyen el G-indicador mejoran la predicción en algunos casos, dependiendo del modelo que se tome como referencia. Permite, no obstante, obtener un indicador ajustado del flujo de turistas británicos con casi un mes de adelanto.

Este es solo un primer paso en la utilización de las búsquedas online para construir indicadores adelantados de actividad económica. Otros temas que podrían explorarse con esta herramienta serían las ventas de automóviles, la confianza de los consumidores y la compra de viviendas. La principal característica de estos procedimientos es que, con el tiempo y los continuos avances en el uso de Internet, los resultados no pueden más que mejorar en el futuro. No obstante es necesario recordar que la construcción de estos G-indicadores ha de ser cuidadosa con el objeto de evitar equívocos derivados, entre otros aspectos, del distinto uso del lenguaje en diferentes países. Ignorar estas cautelas y confiar de forma ciega en estos indicadores que no siempre mejoran la capacidad predictiva de los modelos de previsión a corto plazo más tradicionales puede llevar a la obtención de resultados erróneos.

Palabras clave: Google, predicción, turismo.

Códigos JEL: C22, C53, C82, E17, E37, L83.

Abstract

This paper reviews some of the applications that use the vast swathes of information provided by Internet user searches for economic analysis and forecasting. This enormous volume of information, available in real time, can be handled by analysts thanks to statistical tools such as "Google Insights for Search", which allow trends in different areas of interest to be classified and evaluated. Previous work focused predominantly on the labour market, on the housing market, on retail sales and on consumer confidence. This paper presents a very specific application for the Spanish economy: British tourist inflows to Spain (the Spanish tourist industry's main customers). The improvement in the forecasting provided by the short-term models that include the G-indicator depends on the benchmark model. This does, however, allow an adjusted indicator of the inflow of British tourists to be obtained with a lead of almost one month.

This is but an initial step in the use of on-line searches for constructing leading indicators of economic activity. Other applications to be explored are car sales, consumer confidence and house purchases. The chief characteristic of these procedures is that, with time and the continuous growth of Internet use, results can only improve in the future. It should nonetheless be recalled that the construction of these G-indicators requires caution so as to avoid mistakes arising, *inter alia*, from the different use of language in different countries. Not taking due caution and blindly confiding in these indicators may lead to erroneous results being obtained.

Keywords: Google, forecasting, nowcasting, tourism.

JEL classification: C22, C53, C82, E17, E37, L83.

1 Introducción

La popularización del uso de Internet ha generado toda una serie de nuevas actividades y ha cambiado la manera de realizar actividades tradicionales. Leemos cada vez más la prensa online (llevando a cambiar de forma radical el modelo de negocio tradicional de los periódicos), buscamos información médica en la web antes —y después— de ir al médico, compramos toda una serie de productos y servicios online —libros, música, billetes de avión etc.— participamos crecientemente en redes sociales (Facebook cuenta con cerca de 700 millones de usuarios en el mundo y Twitter con 175 millones de usuarios registrados), escribimos emails, y a través de Internet hacemos gestiones con los bancos y las administraciones públicas, hablamos por teléfono o vemos la televisión. Todo este cúmulo de actividades deja una huella en la Web, generando un ingente volumen de información sobre productos, personas, instituciones, intenciones de compra, intenciones de voto, y un largo etcétera. Esto es lo que se ha denominado **Big Data**, un conjunto de información tan extenso y variado que hace imposible su tratamiento con el software habitual en bases de datos tradicionales.

Esta sobreabundancia de datos tanto en volumen como en áreas de interés está dando lugar a toda una serie de desarrollos. En palabras de Gary King, director del “Harvard’s Institute for Qualitative Social Science”: *“It’s a revolution, we are really just getting under way. But the march of quantification, made possible by enormous new sources of data, will sweep through academia, business and government. There is no area that is going to be untouched”*¹.

La explotación de la masa ingente de información —muchas veces desestructurada— en una forma articulada requiere la utilización masiva de tecnologías de reconocimiento de texto, lo que habitualmente se denomina “sentiment analysis”. Las posibilidades de aplicación incluyen, por mencionar solo algunos ejemplos, el diseño de nuevas estrategias de creación de valor para las empresas², la iniciativa de la reserva Federal de Nueva York de promover una aplicación que permita rastrear las opiniones expresadas en las redes sociales³, o la explotación de su potencial como instrumento de predicción en el análisis de de la salud pública, de la economía del desarrollo⁴ o de la actividad económica.

Este trabajo se enmarca en esa área, siguiendo la línea de trabajos iniciados por Choi & Varian, y basados en “Google Insights for Search” una herramienta estadística diseñada por Google que permite analizar desde distintas ópticas las búsquedas realizadas por los usuarios de Internet. Se trata de un primer paso en el análisis de la ingente información contenida en la Web para la economía española. En este artículo se presenta una aplicación concreta —la entrada de turistas británicos— que ilustra el posible uso para la previsión de variables de demanda y actividad en España. La continua extensión de las actividades vinculadas a la Web hace que los resultados de hoy no puedan más que mejorar la capacidad predictiva en el futuro.

1. The New York Times, 11 de febrero 2012.

2. Véase “Big data: The next Frontier for Innovation, Competition and Productivity”. McKinsey Global Institute, mayo 2011.

3. Sentiment Analysis and Social Media Monitoring Solution RFP. Federal Reserve of New York. En la declaración de intenciones se dice: *“Social media platforms are changing the way organizations are communicating to the public. Conversations are happening all the time and everywhere. There is need for the Communications Group to be timely and proactively aware of the reactions and opinions expressed by the general public as it relates to the Federal Reserve and its actions on a variety of subjects”*.

4. **Global Pulse** es una nueva iniciativa de Naciones Unidas que rastrea las redes sociales con objeto de detectar situaciones de riesgo de distinta índole en los países en desarrollo.

Recuadro 1. Otros usos de la información existente en la Web

En los últimos años ha habido varios desarrollos en distintos ámbitos que tienen en común la explotación de la vasta información disponible en la Web. En este recuadro se recogen aquellos más relevantes de los que tenemos conocimiento.

1. Indicadores de confianza basados en la extracción del sentimiento de los ciudadanos reflejado en los mensajes de **Twitter**. Esta línea de investigación incluye trabajos sobre la predicción de la evolución del mercado de valores en “Twitter mood predicts the stock market”, Bollen, J. et al. (2010) y la construcción de indicadores de confianza —basados en un análisis semántico de los *tweets*— que permiten adelantar el mensaje de los indicadores de confianza convencionales basados en encuestas “From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series” O’Connor, B. et al. (2010). Una referencia más reciente de estas investigaciones se puede encontrar en el Economist (junio 2011) en el artículo “Can Twitter predict the future?”.
2. Otra línea de investigación que explota la información disponible en la Web es el “**Billion Prices Project**” que genera indicadores de inflación diarios a través de la recopilación de precios en las páginas Web de las empresas. El proyecto iniciado en Argentina en 2008 por Cavallo, A. y Rigobon, R., se ha ampliado enormemente dando lugar a la empresa PriceStats, fundada en 2011, que vende informes sobre la evolución de los precios de más de cinco millones de productos en 70 países.
3. Un tercer grupo de proyectos basado en la empresa de software **Recorded Future** rastrea en la Web la información de distinto origen, incluyendo tanto fuentes de noticias tradicionales, como nuevas formas de comunicación —blogs, redes sociales— así como informes gubernamentales. A partir del tratamiento masivo de estos registros asigna probabilidades a futuros acontecimientos a distintos horizontes. Sus productos se venden en su página Web.

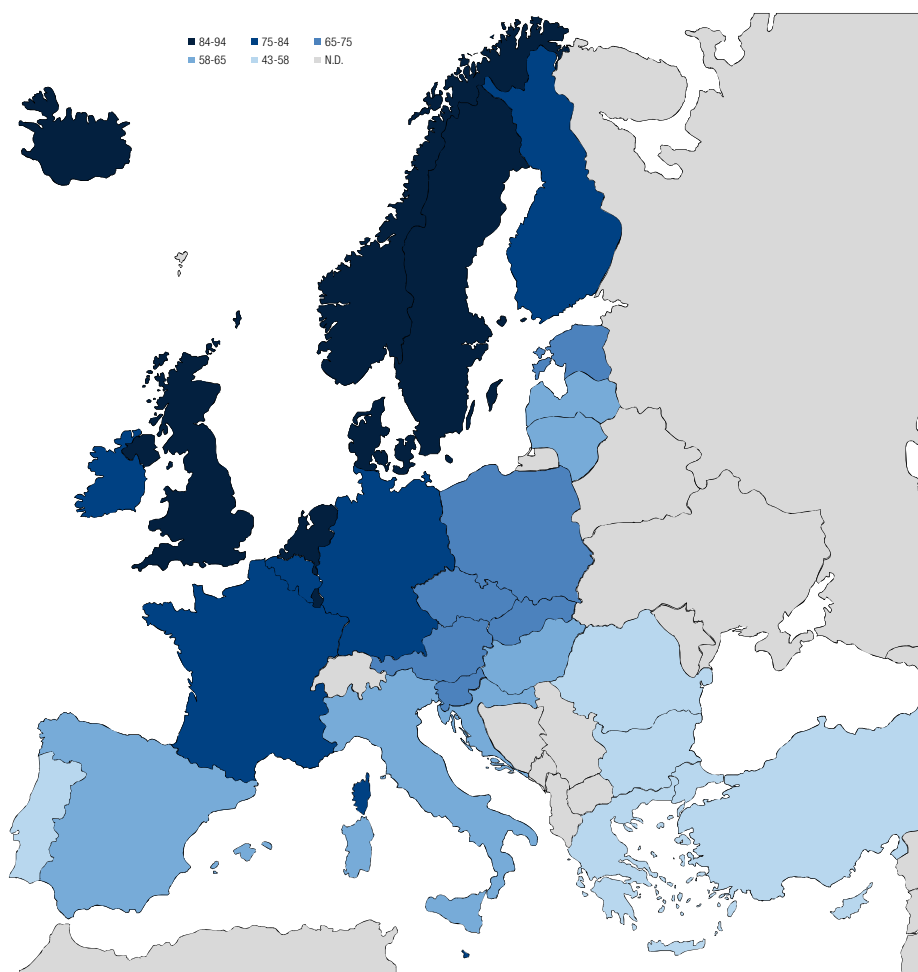
El artículo se organiza de la siguiente forma: en la sección 2 se describen las principales actividades realizadas por los usuarios de Internet, centrándonos en la búsqueda de información de bienes y servicios y cómo estas búsquedas se traducen en compras sea online o en establecimientos tradicionales. En la sección 3 revisamos los trabajos existentes que hacen uso de **Big data** para construir indicadores adelantados, trabajos que cubren áreas muy distintas; desde la detección precoz de la gripe no estacional a las predicciones del precio de la vivienda o de la tasa de paro, pasando por la previsión de ejecuciones hipotecarias. La sección 4 presenta una aplicación para la economía española: la previsión de los flujos de turistas británicos. Por último en la sección 5 presentamos algunas conclusiones y proyectos de trabajo a desarrollar en el futuro.

2 Penetración y usos de Internet en Europa

La importancia de Internet en las actividades cotidianas de la población ha aumentado de forma exponencial en los últimos años. Desde el año 2002, Eurostat lleva a cabo una encuesta anual sobre el acceso y el uso de Internet en los hogares europeos⁵. De acuerdo con la más reciente, correspondiente al año 2011, la fracción de la población adulta (16 a 74 años) sin contacto alguno con Internet (la denominada *exclusión digital*) se ha reducido de forma drástica pasando del 42% en 2006 al 24% en 2011. La agenda digital europea establece como objetivo para 2015 que el nivel de *exclusión digital* no supere el 15% de la población de 16 a 74 años. Algunos países miembros de EU27 cumplen ya holgadamente ese objetivo (en Suecia solo un 5% de la población está excluida del mundo digital, en Holanda un 7%, y en UK el 11%). Centrándonos en los cuatro grandes países continentales de EU27, Alemania con un 16% de excluidos y Francia (18%) están cerca de alcanzar el objetivo fijado para 2015. Sin embargo en Italia y España los hogares excluidos del mundo digital constituyen todavía una proporción abultada de la población: el 39% y el 29% respectivamente en el año 2011.

HOGARES CON ACCESO A INTERNET (EN %) 2011

GRÁFICO 1



FUENTE: Eurostat.

5. "ICT usage in households and by individuals". Eurostat.

En 2011 el 73% de los hogares europeos (EU27) tenían acceso a Internet en contraste con el 41% en 2004. La cobertura de la conexión de banda ancha se ha más que duplicado entre 2006 y 2011 en que ya está disponible para el 68% de los hogares.

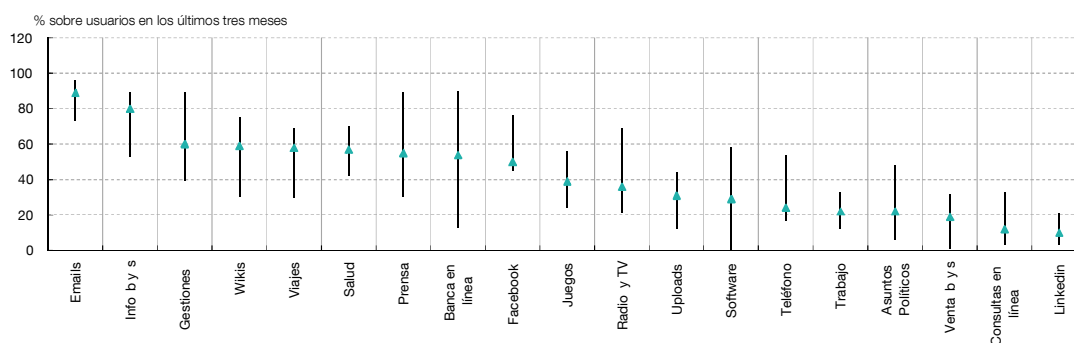
Si bien estos avances se han producido en todos los países, la penetración de Internet en los hogares europeos sigue siendo dispar (véase gráfico 1). Los países más avanzados, con más de un 90% de los hogares conectados a Internet, son Holanda (94%), Dinamarca (90%) y Suecia (91%). En el extremo opuesto, con una cobertura inferior al 50% están Bulgaria, Rumania y Grecia.

2.1 ¿Qué actividades realizan los usuarios?

La encuesta de Eurostat sobre el uso de Internet permite analizar cuáles son las actividades más habituales realizadas por los usuarios (véase gráfico 2). En todos los países el correo electrónico es la actividad más común. El 89% de los usuarios residentes en países del área del euro utilizan Internet para mandar o recibir mensajes. Las diferencias en el uso del correo entre unos países y otros son muy pequeñas. En segundo lugar destaca la búsqueda de información sobre bienes y servicios, seleccionada por el 80% de los encuestados. Un segundo grupo de actividades, realizadas por cerca del 60% de los usuarios, son las vinculadas a gestiones administrativas (60%), consulta de “wikis” para informarse sobre diversos temas (59%), alojamiento y viajes (58%) y búsqueda de información médica (57%).

ACTIVIDADES EN INTERNET (a)

GRÁFICO 2

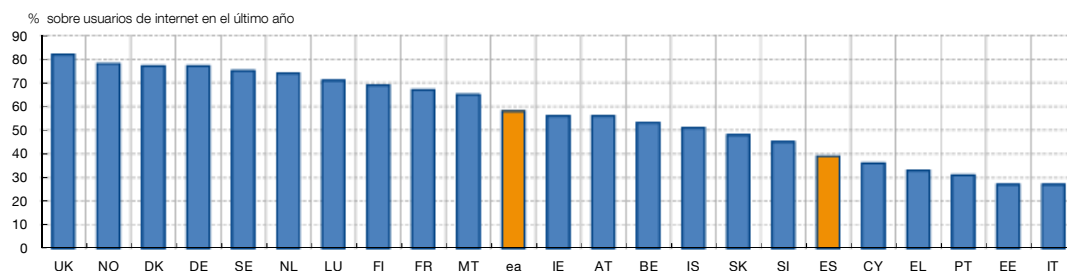


FUENTE: Eurostat.

(a) Los países incluidos son los 17 integrantes de la zona del euro + Dinamarca (DK), Suecia (SE), Noruega (NO), Islandia (IS), Reino Unido (UK).

Las barras verticales reflejan la dispersión de los valores en los 22 países considerados.

En torno al 55% de los usuarios de la eurozona leen la prensa online y operan con su banco a través de Internet. En ambos casos la dispersión entre países es muy grande. Por ejemplo la banca online es utilizada por más del 80% de los usuarios en los países nórdicos (Noruega, Dinamarca, Estonia, Finlandia y Suecia) y solo por un 13% de los usuarios griegos. En el caso de la prensa destacan de nuevo los ciudadanos de los países nórdicos como los más ávidos lectores de prensa online —por encima del 85% de los usuarios— mientras que solo un 30% de los franceses y un 39% de los irlandeses utilizan este canal informativo. La mitad de los usuarios europeos participa en foros sociales (facebook, twitter, etc) siendo su uso bastante uniforme en todos los países y algo superior en los países nórdicos. Para el resto de las actividades la popularidad es muy inferior al 40% y la variabilidad entre países aumenta notablemente.



FUENTE: Eurostat.

La Encuesta de Eurostat incluye un apartado independiente para evaluar la incidencia y características del comercio electrónico. En 2011, el 58% de los usuarios de Internet declaran haber realizado alguna compra online en los doce meses anteriores. Las diferencias entre países son considerables (véase gráfico 3). Destaca el Reino Unido como el país donde la popularidad de las compras online supera el 80% de los usuarios, seguido de Dinamarca y Alemania (77%) y Suecia (75%). En el extremo opuesto están Italia y Grecia donde el comercio electrónico alcanza a menos del 30% de los usuarios. España está por debajo de la media, con un 40% de los usuarios.

2.2 ¿Qué compran los europeos en la Web?

En el cuadro 1⁶ se muestra la distribución de las ventas por Internet en la zona del euro, así como el valor máximo y mínimo y el correspondiente a España. Las compras más habituales en Internet son libros, música, películas y software. En su conjunto, un tercio de los usuarios de Internet ha comprado alguno de estos productos en el año 2010. Un 30% compró viajes o hizo reservas de hoteles, el 27% compró ropa o material deportivo y un 20% compró entradas para algún concierto, teatro, cine, etc.

COMPRAS ONLINE (en % de los usuarios de internet en el último año) (a)

CUADRO 1

	Zona del euro	MAX	MIN	ES
Películas / música o libros, revistas, material de aprendizaje, software	33	56	9	13
Viajes y alojamiento de vacaciones	30	53	9	23
Ropa, artículos deportivos	27	68	7	10
Entradas para eventos	20	51	2	15
Artículos del hogar	19	34	2	7
Películas / música o libros, revistas, software, entregado o actualizado on line	16	33	3	7
Equipos electrónicos	14	27	4	7
Hardware	12	19	2	6
Productos farmacéuticos	7	20	0	0
Alimentos / comestibles	6	19	1	4
Otros	6	26	1	6
Acciones, servicios financieros o seguros	4	16	1	2

FUENTE: Eurostat y Banco de España.

a. Los países incluidos son los señalados en el Gráfico 2.

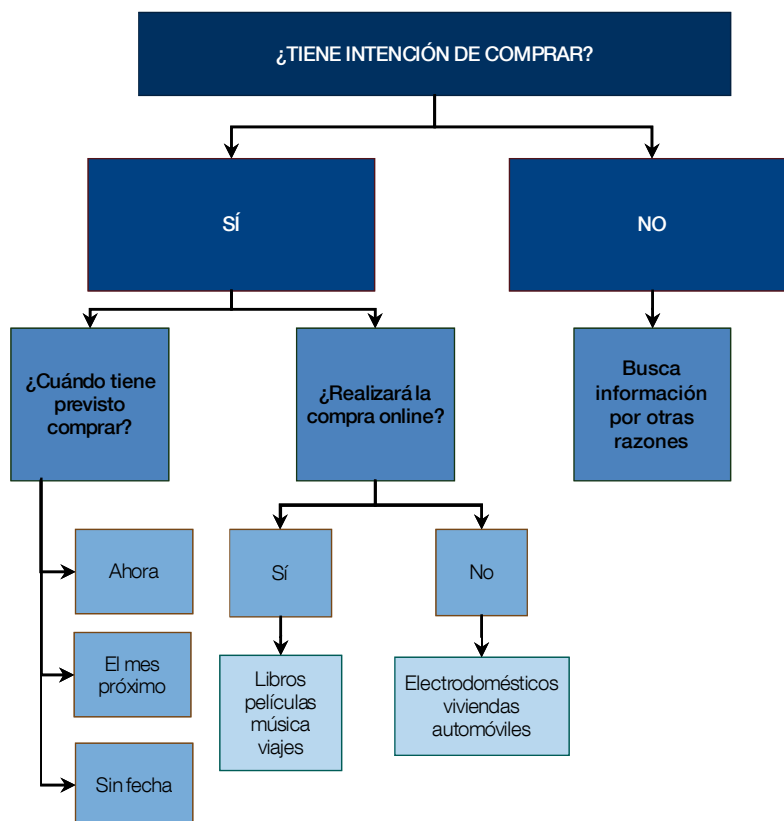
6. La encuesta de Eurostat permite evidentemente las respuestas simultaneas, es decir cada entrevistado debe seleccionar todas las categorías de compra que haya realizado en la Web.

2.3 ¿En qué medida las búsquedas se materializan en compras?

Dado que una de las actividades favoritas de los usuarios es la búsqueda de información sobre bienes y servicios, cabe preguntarse en qué medida estas búsquedas de información se corresponden con decisiones de compra. Cabría esperar que parte de las búsquedas sobre un determinado producto tenga el propósito de recabar información de cara a la realización de la compra. En algunos casos esa compra se realizará prácticamente de forma simultánea a la búsqueda de la información (compra de música o de libros), incluso no es seguro que una compra online tenga que venir precedida necesariamente de una búsqueda en la Web (la compra de entradas). En otros casos, especialmente cuando se trata de compras de bienes de consumo duradero de cierta importancia (compra de viviendas, de automóviles o electrodomésticos) el comportamiento más típico es realizar una búsqueda online, comparando productos, precios y proveedores para, posteriormente, realizar la compra en un establecimiento convencional. No parece que sea muy común comprar un automóvil online aunque solo fuera por la necesidad de probarlo antes, sin embargo, si es muy habitual la búsqueda de información en la Web semanas o meses antes de realizar su compra. El desfase temporal entre las búsquedas y las compras abre la posibilidad de construir un indicador adelantado de ventas de automóviles. En el gráfico 4 se resumen los distintos casos posibles.

SE BUSCA INFORMACIÓN SOBRE BIENES Y SERVICIOS EN LA WEB

GRÁFICO 4

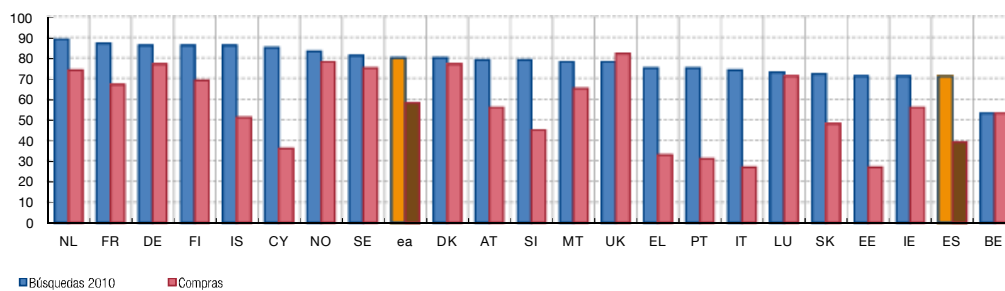


FUENTE: Elaboración propia.

A nivel agregado, la relación entre compras y búsquedas de bienes y servicios no es muy elevada (véase gráfico 5): mientras que —tal y como se mencionaba en la sección anterior— la preferencia por buscar información sobre productos está bastante extendida en los distintos países, existen grandes discrepancias entre países en la propensión a realizar compras online.

BÚSQUEDAS Y COMPRAS EN LA WEB (a)

GRÁFICO 5



FUENTE: Eurostat (Encuesta 2010).

(a) Las búsquedas se miden en % de usuarios de internet en los últimos tres meses precedentes. Las compras se refieren al porcentaje de usuarios en el último año.

3 Construcción de indicadores adelantados a partir de las búsquedas realizadas por los usuarios

El lanzamiento de la herramienta denominada Google Insights for search (GIS)⁷ hace posible utilizar la ingente información sobre las búsquedas realizadas por los usuarios de Internet para construir indicadores adelantados en distintas áreas de conocimiento. En el recuadro 2 se resumen las ideas básicas de esta herramienta; se puede encontrar una descripción más detallada en <http://www.google.com/insights/search/#>

Recuadro 2. Ideas básicas de Google Insights for Search (GIS)

En este recuadro se presentan las ideas básicas de la herramienta Google Insights for Search (GIS).

- GIS analiza una **muestra** aleatoria de las búsquedas realizadas por los usuarios y no el total de búsquedas.
- Los índices generados por GIS representan la popularidad de las búsquedas de un término en relación con el total de búsquedas realizadas en un determinado ámbito geográfico y lapso temporal medida en términos relativos. Las series generadas por GIS no proporcionan por tanto valores absolutos del volumen de búsquedas, solo la frecuencia relativa de las mismas. Por ejemplo, si se introduce el término de búsqueda **té**, con el parámetro de ubicación definido en **Escocia** y el parámetro de tiempo en **marzo de 2007**, el indicador GIS nos proporcionará la popularidad de este término entre los usuarios de Escocia en marzo de 2007 —para una muestra de búsquedas— en comparación con las búsquedas de todo tipo de términos realizadas en Escocia en marzo de 2007.
- Los índices están **normalizados en una escala de 0 a 100** dividiendo la popularidad relativa en cada momento *t* por el máximo en el período de referencia. La frecuencia es semanal.
- Para preservar la calidad de los indicadores, GIS fija un **umbral de tráfico mínimo** por debajo del cual el índice se fija en 0. Además para evitar que se puedan manipular artificialmente por los usuarios, GIS **elimina las búsquedas repetidas** realizadas por un mismo usuario en un corto lapso de tiempo.
- El ámbito de búsqueda se puede acotar en función de varios parámetros: **Zona geográfica**, **período temporal** (desde 2004) y **categorías de búsqueda** solo para algunos idiomas (inglés, alemán, italiano, japonés, ruso, sueco y polaco). Desde diciembre de 2011 la categorías están disponibles también en español, árabe, esloveno y turco entre otros.
- Las búsquedas se clasifican en categorías que obedecen a asignación de las búsquedas a industrias o mercados determinados. Hay **27 categorías** denominadas de primer nivel y **241 subcategorías**. Esta clasificación se realiza de forma automática a partir de un procesador de lenguaje natural (*semantic orientation device*). Por ejemplo la categoría Automóviles y vehículos incluye hasta 21

7. En alguna documentación se refieren a esta herramienta como Google Trends, su denominación inicial. Google Insights for search es una versión más completa de la anterior.

subcategorías, entre la cuales están las denominadas: Industria del automóvil, bicicletas y accesorios, barcos, motocicletas, ferias de automóviles.

- GIS está a disposición de los usuarios y es gratis.

- **Algunas cautelas a tener en cuenta.**

Cada vez que se utiliza GIS para analizar la popularidad relativa de un término, GIS selecciona una muestra aleatoria nueva. La nueva muestra se guarda en los servidores de Google durante un día; esto hace que el Índice pueda variar de un día a otro, aunque no parece que las variaciones sean demasiado importantes.

Es importante contrastar que el índice de búsqueda está proporcionando la información que realmente se indaga, no se debe utilizar la herramienta de una manera puramente automática a riesgo de obtener resultados disparatados. Por ejemplo, supongamos que tenemos interés en saber cuales los aquellos países donde se genera un interés relativamente importante por viajar a España. Se buscaría la palabra SPAIN en la categoría de viajes. ¿Dónde se produce el mayor interés relativo en viajar a España? De acuerdo con GIS obtendríamos Irlanda como el país donde una proporción mayor de las búsquedas en Internet se refieren a viajes a España (normalizada al valor 100 para la media 2004-hoy), seguido de Reino Unido con un valor de 84. El tercer país es Trinidad y Tobago, con un valor 64, lo cual parece sorprendente por un momento, antes de darnos cuenta que la capital de Trinidad y Tobago —uno de los pocos países del Caribe de habla inglesa— es Port of Spain.

3.1 Los trabajos pioneros: gripe y paro

Uno de los trabajos pioneros se enmarca en el campo de la epidemiología. Ginsberg, J. et al. (2008) analizaron los millones de búsquedas acerca de brotes de gripe de carácter no estacional y su relación con las visitas al médico. En el trabajo estiman con bastante precisión el nivel de incidencia semanal de la gripe en cada región de los Estados Unidos y generan indicadores que con un solo un día de retraso, adelantan la información estadística más convencional permitiendo un tratamiento más temprano y eficaz de la enfermedad.

Entre las aplicaciones precursoras en el campo de la economía están los trabajos de Choi y Varian “Predicting the Present with Google trends” y “Predicting Initial Claims for Unemployment Benefits”, ambos publicados en 2009 en el Google Research blog. Las búsquedas en Google realizadas por los usuarios permiten derivar información en tiempo real sobre el comportamiento de los consumidores en distintos mercados. Se analiza en concreto como Google Trends permite mejorar las predicciones a corto plazo en **ventas minoristas, ventas de automóviles, actividad en mercado inmobiliario y viajes**. En muchos casos la inclusión de las series de los índices generados por Google Trends permite mejorar la predicción obtenida con modelos convencionales. La estrategia en cada una de estas aplicaciones es la siguiente:

- Ajustar el mejor modelo de predicción posible utilizando las estadísticas habituales, incluyendo la variable endógena desfasada (Modelo 0 en los trabajos referidos).
- Añadir los índices de *Google Trends* como una variable explicativa adicional (Modelo 1 en estos mismos trabajos).

- Evaluar la mejora en las predicciones. Típicamente a través del error absoluto medio (EAM) de las predicciones extramuestrales utilizando una “rolling window forecast”.

¿En qué medida este procedimiento permite mejorar las predicciones de los modelos de corto plazo convencionales?

Los resultados varían de unos casos a otros. En algunos la mejora es pequeña mientras que en otros es muy sustancial. Por ejemplo en el caso de las ventas de automóviles las ganancias obtenidas al incluir un indicador de búsquedas en Google es de un 18% en relación con el modelo convencional. En contraste, en el mismo trabajo Choi & Varian presentan los resultados de incluir las búsquedas en Internet para obtener predicciones más ajustadas de las ventas de vehículos Ford. Los resultados son, en este caso, muy pobres, obteniéndose tan solo una mejora del 3% en relación con el modelo alternativo. Con una metodología similar, Choi & Varian estiman una mejora del 15% en las previsiones de demandantes de subsidio de desempleo cuando incorporan un indicador basado en las búsquedas en Google de términos relacionados con el desempleo.

3.2 Otros campos de aplicación: mercados de la vivienda, tasa de paro, confianza de los consumidores

Con posterioridad se ha publicado un buen número de trabajos utilizando esta herramienta en distintas actividades y zonas geográficas. Chamberlin (2010) explora la posibilidad de construir indicadores de ventas minoristas, transacciones inmobiliarias, matriculaciones de automóviles y viajes al extranjero a partir de las búsquedas online realizadas en el Reino Unido. Bughin (2011) realiza un ejercicio similar para la economía belga obteniendo que un 10% de aumento en la intensidad de las búsquedas de términos relacionados con compras se traduce en un aumento del 2% en las cifras de ventas en el trimestre de referencia. En el caso del desempleo un aumento del 10% de las búsquedas adelanta un aumento del 1% en el número de demandantes del subsidio de paro.

Una de las líneas de trabajo más común es la de utilizar esta técnica para mejorar la predicción del desempleo. D’Amuri & Marcucci (2009) encuentran una enorme mejora en las previsiones de la tasa de paro en Estados Unidos cuando se incluye un indicador basado en las búsquedas en Internet. Askitas & Zimmermann (2009) obtienen que las predicciones de la tasa de paro en Alemania son mejores cuando se incluyen en el modelo las búsquedas de trabajo realizadas online en las dos últimas semanas del mes anterior. Fondeur & Karamé (2011) estiman ganancias en las previsiones de desempleo en Francia cuando se incorporan las búsquedas online que dependen del grupo de edad: la precisión de las previsiones aumenta en torno a un 20% para los jóvenes de 15 a 24 años, la mejora es de un 18% en el grupo de 25 a 49 años mientras que para los mayores de 50 años la evidencia estadística no permite obtener un resultado concluyente.

Wu & Brynjolfsson (2009) y Webb (2009) aplican los indicadores generados por Google al mercado de la vivienda en Estados Unidos. El primero de estos trabajos se centra en la predicción de precios y volúmenes de venta en el mercado de la vivienda, y el segundo en el mercado hipotecario analizando la fuerte correlación entre la palabra clave “foreclosure” (desahucio) y la estadística de desahucios. Una línea de trabajo que empieza a ser habitual es la construcción de indicadores de confianza. Della Penna y Huang (2009) construyen un indicador de confianza de los consumidores en base a la popularidad relativa de determinadas búsquedas (como quiebra, mobiliario de oficina o bienes de lujo) que permite

adelantar la tendencia de los indicadores de confianza convencionales basados en encuestas y supera a estos en la predicción del consumo privado de la economía americana.

En definitiva, en la medida en que la realidad de un país se refleja en las búsquedas online realizadas por sus ciudadanos, la información agregada de estas búsquedas permite construir toda una serie de indicadores en tiempo real de la coyuntura económica presente y futura, lo que puede resultar útil también para bancos centrales a la hora de hacer previsión económica. Esta es la idea que subyace en el reciente trabajo publicado por el Banco de Inglaterra⁸ donde, tras explorar el potencial de los indicadores basados en búsquedas para el mercado de trabajo y los mercados de vivienda en el Reino Unido —con mejores resultados en el segundo caso— concluyen afirmando que el Banco: *“will continue to monitor these data as part of the range of different indicators it considers in forming its view about the Outlook for the economy of the UK”*.

Más recientemente, la Reserva Federal de Nueva York ha publicado en su blog un artículo “Forecasting with Internet Search data” donde se exploran las posibilidades de aplicar las búsquedas realizadas por los usuarios de Internet a la evolución presente y futura de los mercados financieros, en concreto al mercado hipotecario y a las expectativas de apreciación/depreciación del renminbi.

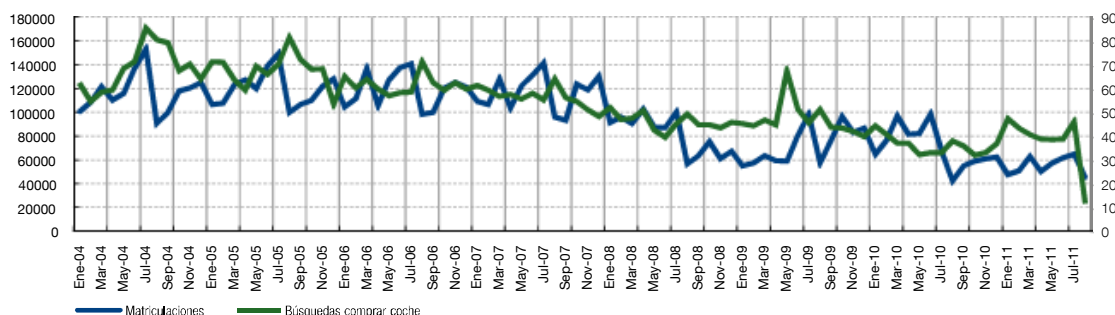
8. “Using Internet search data as economic indicators” QB 2011Q2.

4 Una aplicación para la economía española: Los flujos de turistas

Al plantearnos la utilización de esta herramienta para la economía española, surgieron varias opciones; en concreto el mercado de trabajo parecía un campo de investigación prometedor donde, además, podíamos seguir la estela de los trabajos realizados en otros países. Por otra parte, siguiendo el esquema del gráfico 4, *Google Insights for Search (GIS)* permite la construcción de indicadores adelantados de la adquisición de bienes de consumo duradero y de algunos servicios, en la medida en que esas compras vienen precedidas en muchas ocasiones de búsquedas en Internet⁹. Por ejemplo, se puede analizar la relación entre las búsquedas del término “comprar coche” y las matriculaciones de automóviles por particulares (véase gráfico 6). Sin embargo, fueron tres las razones que nos llevaron a explorar el potencial de esta herramienta en la predicción de los flujos de turistas extranjeros. La primera de carácter puramente técnico; cuando se inició este proyecto —julio de 2011— la categorización de las búsquedas no estaba disponible en español (véase recuadro 3)¹⁰, lo que nos llevó a buscar una aplicación en la que la inexistencia de las categorías en castellano no fuese un obstáculo. De ahí surgió la idea de analizar los flujos de turistas extranjeros. Los idiomas de búsqueda más relevantes serían inglés, alemán y francés, idiomas para los que GIS sí disponía de la categorización de las búsquedas. La aplicación al turismo de las búsquedas online era un campo casi inexplorado, con la excepción del trabajo de Choi & Varian mencionado más arriba¹¹.

BÚSQUEDAS DE COMPRA DE COCHES ONLINE

GRÁFICO 6



FUENTES: ANFAC y Google Insights for Search.

En segundo lugar, el turismo es una de las actividades donde se realiza una utilización más intensiva de las búsquedas por Internet, tanto para obtener información anticipada sobre destinos y paquetes, como para la adquisición online de esos viajes, por lo que la estadística de búsquedas podría proporcionar datos muy valiosos —y posiblemente adelantados— sobre la trayectoria futura de los flujos turísticos.

9. También es un instrumento útil para detectar la evolución en tiempo real de otros productos de consumo donde no hay un desfase temporal entre su búsqueda y su adquisición, por ejemplo libros, música etc.

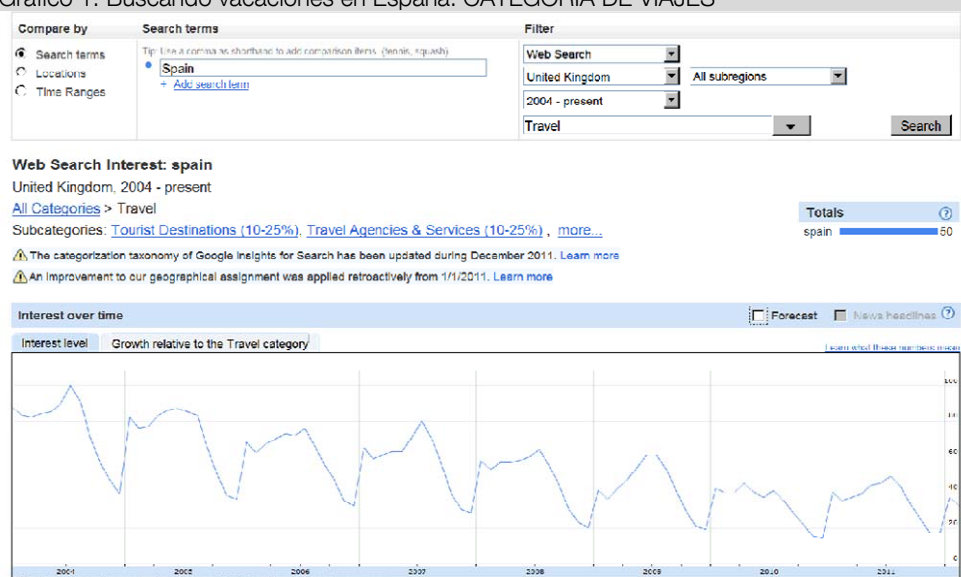
10. Las categorías no son imprescindibles aunque facilitan enormemente las primeras fases del análisis.

11. En enero 2012 BBVA presentó un indicador de turismo basado también en Google. Por el momento no se encuentra disponible la información relativa a los modelos utilizados.

Recuadro 3. Búsquedas basadas en categorías

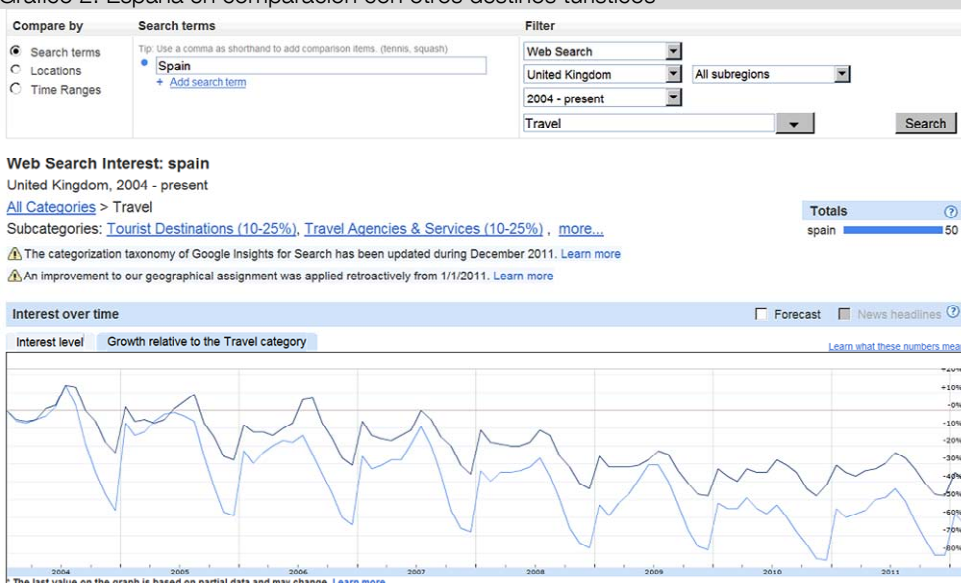
Alternativamente se puede hacer uso de la categorización automática proporcionada por GIS y buscar el término Spain dentro de la categoría de viajes (véase gráfico 1). Se confirma la tendencia decreciente de las búsquedas, más acusada que cuando las búsqueda se hacen en base a la búsqueda manual de términos relacionados con “holiday Spain”. GIS genera una información adicional cuando las búsquedas se realizan dentro de una categoría determinada: proporciona la evolución del interés en el término de búsqueda en relación con su categoría. En este caso compara el interés relativo en viajar a España con el de viajar a cualquier otro lugar (véase gráfico 2). La popularidad de las búsquedas de viajes ha caído un 26% desde 2004, la popularidad relativa de los viajes a España ha caído en torno al 50%.

Gráfico 1. Buscando vacaciones en España. CATEGORIA DE VIAJES



Fuente: Google Insights for Search.

Gráfico 2. España en comparación con otros destinos turísticos



Fuente: Google Insights for Search.

Finalmente, el interés de analizar los flujos de turismo se justifica plenamente dada la gran contribución de la industria turística a la economía nacional. España no solo es uno de los principales destinos turísticos del mundo —el cuarto en términos de número de turistas y el segundo en términos de ingresos por turismo— sino que además es el país desarrollado con mayor dependencia del turismo en el mundo (véase cuadro 2).

EL TURISMO EN ESPAÑA, EN PERSPECTIVA

CUADRO 2

	Llegadas internacionales (millones) 2010	Ingresos por turismo (mmde \$) 2010	Turismo en % del PIB 2009	Turismo en % del empleo 2009
Francia	76,8	46,3	3,7	na
Estados Unidos	59,7	103,5	2,6	na
China	55,7	45,8		
España	52,7	52,5	10,7	12,7
Italia	43,6	38,8	4,8	9,7
Reino Unido	28,1	30,4	2,7	na
Turquía	27,0	20,8		
Alemania	26,9	34,7	3,2	na
Malasia	24,6			
Méjico	22,4			

FUENTES: Organización Mundial del Turismo (2011) y OCDE (2010).

Los principales clientes de la industria turística española son Reino Unido, Alemania y Francia (véase cuadro 3). Desde el máximo alcanzado en 2007 los turistas entrados han disminuido en un 3%. Los turistas británicos son los que han sufrido un mayor descenso desde 2007 (16%). La reducción en los viajeros procedentes de Alemania y Francia ha sido muy inferior, del 10% y 6% respectivamente. El número de viajeros procedente de otros países europeos también se ha contraído en mayor o menor medida, con la excepción de los países nórdicos —Dinamarca, Finlandia, Noruega y Suecia— cuyos turistas con destino a España han aumentado un 13% en el período, y los procedentes de Rusia que han más que duplicado su número entre 2007 y 2010 si bien su dimensión es aún escasa (850.000 personas en 2011). El turismo procedente del resto del mundo movió a 4,8 millones de personas a España en 2011, compensando en parte el declive en los flujos turísticos procedentes de la mayor parte de los países europeos.

ENTRADAS DE TURISTAS SEGÚN PAÍS DE RESIDENCIA

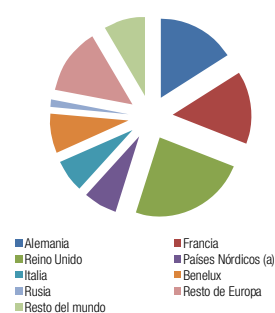
CUADRO 3

	2007	2011	2011	2011/2007
Alemania	10,1	9,0	17%	-10%
Francia	9,0	8,5	16%	-6%
Reino Unido	16,3	13,6	26%	-16%
Países Nórdicos (a)	3,4	3,9	7%	13%
Italia	3,6	3,8	7%	4%
Benelux	4,2	4,6	9%	8%
Rusia	0,4	0,9	2%	101%
Resto de Europa	7,8	7,7	15%	1%
Resto del mundo	3,7	4,8	9%	26%
DE+UK+FR	35,4	31,1	59%	-12%
Total	58,7	56,7	100%	-3%

FUENTE: IET.

a. Países Nórdicos: incluye Dinamarca, Finlandia, Noruega y Suecia.

Turistas 2011

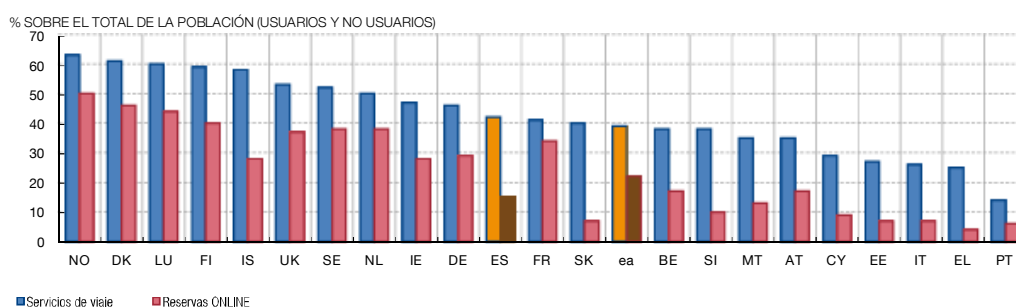


4.1 Aplicando Google Insights for search (GIS) a los flujos de turistas

De acuerdo con la encuesta sobre el uso por familias e individuos de Tecnologías de la Comunicación y de la Información (TIC)¹², un 39% de los residentes en la eurozona llevan a cabo actividades online relacionadas con viajes, mientras que un 22% hacen compras online de este tipo de servicios. La brecha es por tanto de 17 pp¹³. El tamaño de esta discrepancia varía mucho de unos países a otros (véase gráfico 7), aunque las dos estadísticas proporcionan un mensaje similar: la correlación entre ambas es 0,903. Las mayores desviaciones se observan en Islandia —uno de los países con mayor implantación de Internet— donde más del 58% de los ciudadanos se informan pero apenas el 28% compra viajes online. Otros países donde la brecha es relativamente grande son Eslovenia, Eslovaquia y España. Las razones que explican la mayor discrepancia en estos países podría obedecer a la escasa confianza de los usuarios en la seguridad en la realización de compras online o a un menor desarrollo de la venta online por parte de las empresas que operan en la Unión Europea, o quizás a ambos, ya que ambos obstáculos se retroalimentan.

SERVICIOS DE VIAJE Y RESERVAS ONLINE (a)

GRÁFICO 7



FUENTE: Eurostat (Encuesta 2010).

(a) Los servicios de viaje incluyen el uso de internet para la recopilación de información y la adquisición de bienes y servicios relacionados con alojamiento de viajes / vacaciones. Las reservas online incluyen aquellas realizadas en los últimos doce meses.

Desde el año 2009 la Comisión Europea elabora una encuesta específica sobre turismo —“Survey on the attitudes of Europeans towards tourism”— que recaba información detallada sobre usos y costumbres en los viajes de los europeos incluyendo: el método de transporte usado en sus vacaciones, los aspectos financieros asociados, sus destinos favoritos, sus planes para las próximas vacaciones o las razones por las que no tomaron vacaciones el años anterior¹⁴. Según esta encuesta, el porcentaje de la población que consulta Internet como primera fuente de información turística ha avanzado en 6 pp, desde el

12. “EU Survey on ICT usage in households and by individuals”. La encuesta se dirige a hogares donde al menos una persona esté entre 16 y 74 años y a los individuos de ese mismo rango de edad. Se pregunta a los hogares sobre el acceso a Internet de cualquier miembro de la unidad familiar dentro del domicilio de residencia. Las preguntas a los individuos incluyen la frecuencia y el lugar de acceso, las actividades realizadas y su habilidad para realizarlas. La muestra para EU27 en la edición 2011 de esta encuesta es de cerca de 150.000 hogares y más de 200.000 individuos.

13. No se sugiere que el segundo sea un subconjunto del primero, tan solo se comparan los tamaños de los dos grupos.

14. Las entrevistas son telefónicas en la mayoría de los casos. La muestra es de 30.000 personas con más de 15 años y que residen en países de EU27, Croacia, Turquía, Macedonia, Noruega e Islandia.

38% en 2009 al 45% en 2011¹⁵, situándose solo en segundo lugar tras las recomendaciones de amigos y colegas —fuente favorita para 58% de los europeos— y muy por delante de la experiencia personal (21%) y las agencias de viaje tradicionales (21%)¹⁶. En definitiva, casi la mitad de los ciudadanos europeos buscan información sobre viajes y vacaciones prioritariamente en Internet. Esta vasta información sobre potenciales viajes es el sustrato sobre el que se propone compilar un indicador del interés en realizar viajes a un lugar determinado, realizando una aplicación específica para los viajes a España.

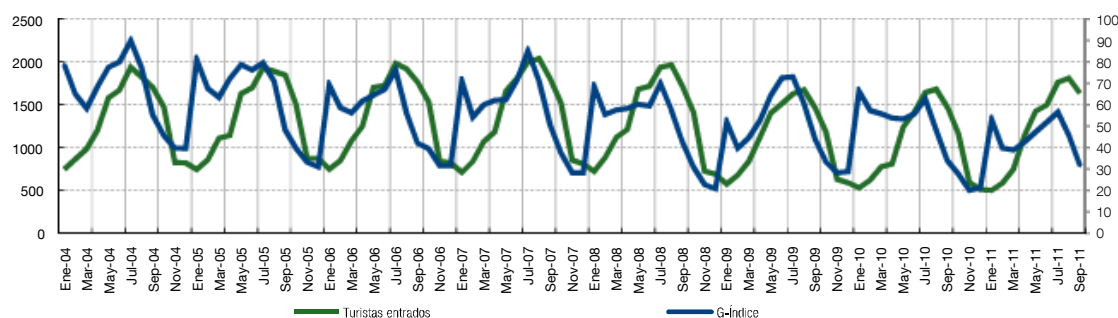
Para ilustrar sobre el potencial de esta herramienta pero también sobre las cautelas con las que se debe utilizar, la discusión que se realiza a continuación se centra en el subgrupo de viajeros británicos que buscan vacaciones en España. Los británicos constituyen los principales clientes de la industria turística española, y además están en el grupo de cabeza en el uso de Internet en Europa, por lo que resulta un caso idóneo para utilizarse como ejemplo en la construcción de un indicador adelantado de los flujos de turistas.

El punto de partida son las búsquedas realizadas desde el Reino Unido del término “Spain holiday”¹⁷. El interés relativo de los británicos ha descendido de forma casi continua desde 2004. En julio de 2011 el indicador de búsquedas alcanzaba un valor de 63, en relación con el máximo de 100 en julio de 2004, una caída del 37%. Para interpretar correctamente estas cifras hay que recordar que se trata de ratios relacionando las búsquedas del término “Spain holiday” con el total de búsquedas de todo tipo realizadas por los británicos. En definitiva este declive mostrado refleja simplemente que todas las demás búsquedas han crecido más que las búsquedas de vacaciones en España.

Tal y como se ha señalado más arriba las entradas en España de turistas británicos han disminuido un 9% entre 2004 y 2011. El perfil de evolución de ambas series no es muy diferente (véase gráfico 8) y el G-index parece adelantar en un par de meses las llegadas de turistas.

TURISTAS BRITÁNICOS Y BÚSQUEDAS EN GOOGLE

GRÁFICO 8



FUENTES: IET y Google Insights for Search.

15. Nótese que estas cifras son coherentes con la Encuesta sobre el uso por familias e individuos de Tecnologías de la Comunicación y de la Información (TIC).

16. Otras fuentes de información son mucho menos populares: los folletos no comerciales (13%) —por ejemplo de oficinas oficiales de turismo—, las guías de viajes (13%) o los *media* convencionales.

17. Alternativamente se puede hacer el ejercicio incluyendo el término Spain y buscando únicamente en la categoría de viajes. Los resultados, muy similares, se presentan en el recuadro 4.

La elevada correlación entre ambas series refleja en buena medida su fuerte componente estacional, aunque persiste una correlación razonable cuando estas se calculan sobre las series ajustadas de estacionalidad. Las mayores entradas de turistas británicos se producen en los meses de julio y agosto, el flujo decrece ligeramente en septiembre, algo más en octubre, y alcanza su mínimo en los meses de noviembre a enero. Este carácter fuertemente estacional parece haberse agudizado en los últimos años como resultado de una mayor caída en el turismo fuera de la temporada julio-agosto¹⁸. Esta mayor concentración de las llegadas de turistas en los meses centrales del verano parece causada por la persistencia de la crisis económica que ha llevado a los británicos a reducir sus viajes de vacaciones al extranjero. Por su parte las búsquedas en Internet de viajes a España reflejadas en el indicador de Google muestran también una estacionalidad muy marcada, siendo julio y enero los meses de mayor dinamismo en las búsquedas de vacaciones. Noviembre y diciembre son los meses de menor actividad. Las pautas estacionales son bastante similares, con la excepción del mes de enero donde la entrada de turistas es mínima aunque las búsquedas de vacaciones en la Web son importantes. Las correlaciones (véase cuadro 4) se calculan para la serie de turistas británicos y el G-indicador contemporáneo y con hasta seis retardos¹⁹. Con las series brutas la correlación más elevada se produce con el retardo 2, es decir entre las búsquedas en un mes y los viajeros entrados dos meses después, sin embargo para las series desestacionalizadas la máxima correlación es la contemporánea.

CORRELACIONES ENTRE TURISTAS BRITÁNICOS Y EL G-ÍNDICE DE BÚSQUEDAS (a)

CUADRO 4

	Desfases del G-Índice						
	0	1	2	3	4	5	6
Datos originales	0,42	0,67	0,72	0,56	0,31	0,00	-0,25
En tasas	0,00	0,32	0,40	0,22	0,31	-0,12	-0,06
En tasas interanuales	0,41	0,32	0,22	0,33	0,10	0,19	0,17
Ajustados de estacionalidad	0,61	0,59	0,57	0,58	0,59	0,61	0,63

FUENTES: IET y Google Insights for Search.

a. Series ajustadas de estacionalidad con Tramo-Seats.

En el cuadro 5 se resumen los resultados de las estimaciones realizadas. En el panel 1 se compara un modelo ARIMA muy simple que normalmente presenta buenas propiedades en la realización de predicciones con el mismo modelo aumentado con el G-indicador contemporáneo²⁰ para la serie de turistas británicos. La media de los errores de predicción mejora en un 18%, lo cual no es despreciable. No obstante, esta ganancia se cancela cuando la comparación se efectúa con un segundo modelo ARIMA más sintético en el que el componente regular, no significativo en la primera estimación, se hace nulo. En este caso el error medio no es significativamente distinto al obtenido en el modelo ARIMA sintético seleccionado por TRAMO.

18. La ratio Max/min, medida por el cociente entre viajeros llegados en julio o agosto y los llegados en enero pasa de 2,5 en los primeros años a 3,5 en los dos últimos.

19. Las series de búsquedas proporcionadas por Google son semanales. Se han periodificado en términos mensuales teniendo en cuenta el número de semanas de cada mes para poder compararlas con las series mensuales de turistas.

20. Inicialmente se incluyó también el G-indicador con hasta seis retardos de desfase, sin embargo ninguno de ellos resultó significativo.

Panel 1: Modelo MA(1)*MA(12) y MA(1)*MA(12) aumentado con Google							
Modelos estimados por Máxima Verosimilitud.	θ	ϕ		Error estándar de los residuos	AIC	BIC	ESM fuera de la muestra
Est	-0,12	-0,40		51,8	509,4	8,1	0.5131 E+04
SE	0,15	0,19					
Panel 2: Modelo MA(12) y MA(12) aumentado con Google							
	θ	ϕ	G	Error estándar de los residuos	AIC	BIC	ESM fuera de la muestra
Est	-0,17	-0,28	10,20	49,4	504,8	8,1	0,3961 E+04
SE	0,16	0,18	3,83				
Panel 3: Modelo MA(12) y MA(12) aumentado con Google							
	θ	ϕ	G	Error estándar de los residuos	AIC	BIC	ESM fuera de la muestra
Est	0,00	-0,38		51,7	508,1	8,0	0,3588 E+0,4
SE		0,13					
Panel 4: Modelo MA(12) y MA(12) aumentado con Google							
	θ	ϕ	G	Error estándar de los residuos	AIC	BIC	ESM fuera de la muestra
Est	0,00	-0,17	13,08	45,0	495,3	7,9	0,3738 E+04
SE		0,14	3,53				

FUENTE: Banco de España.

a. Los modelos se obtienen por TRAMO que corrige un outlier aditivo en Abril de 2009.

b. Los modelos se han estimado por TRAMO para el período 2006 a 2010 (60 observaciones). Con los parámetros obtenidos en esas estimaciones, se realizan predicciones un período por delante para el año 2011.

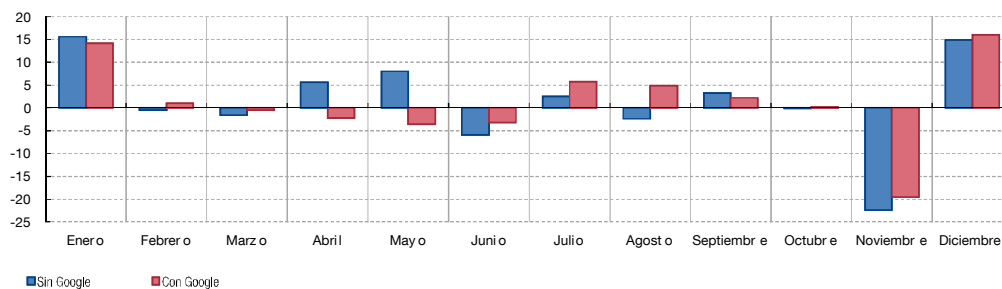
En definitiva, la mejora en la predicción con respecto a los modelos convencionales depende crucialmente de cuál sea el modelo ARIMA que se tome como *benchmark*. Es esta una consideración a tener en cuenta al presentar los beneficios para la previsión derivados de la utilización de los G-indicadores. No obstante, esto no implica que la información contenida en las búsquedas de Google no sea valiosa para la realización de predicciones (véase gráfico 9), sino tan solo que la potencia de estos indicadores ha de juzgarse en el contexto adecuado.

En todo caso, cabe recordar que estos ejercicios se han realizado para un período de estimación todavía corto²¹. En principio al incrementar el número de observaciones, las estimaciones del modelo aumentado con Google deberían mejorar en relación con las del modelo de previsión ARIMA. Por otra parte, los G-indicadores proporcionan información en tiempo real. Por ejemplo, el dato de turistas correspondiente al mes de mayo no se publica hasta finales de junio, mientras que el indicador de Google está disponible el último día de mayo, por lo que el indicador construido a partir de las búsquedas permite mejorar la información disponible para la realización del análisis de coyuntura.

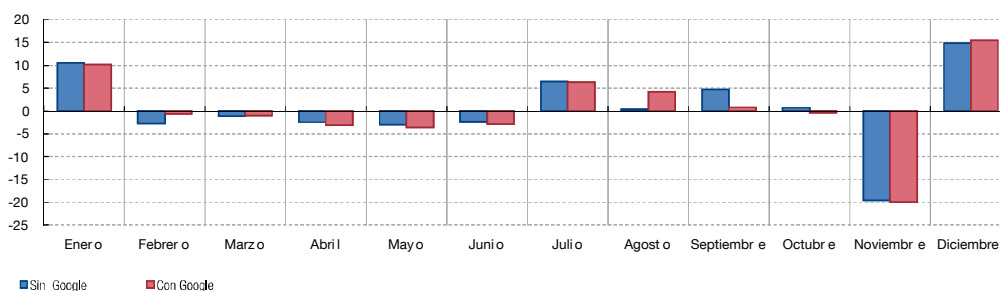
ERRORES DE PREVISIÓN UN PERÍODO POR DELANTE

GRÁFICO 9

Panel 1



Panel 2



FUENTES: IET, Google Insights for Search y Elaboración propia.

²¹. Todos los modelos se han estimado para el período 2006-2010 (60 observaciones) y posteriormente. Las estimaciones un período por delante se han realizado con los parámetros fijados en esas estimaciones.

5 Conclusiones y desarrollos futuros

En este trabajo hemos revisado algunas de las aplicaciones desarrolladas en el campo del análisis y la previsión económica que utilizan la vasta información proporcionada por las búsquedas realizadas por los usuarios de Internet. Esta ingente cantidad de información, muy rica en volumen y obtenida en tiempo real, es manejable para los analistas gracias a herramientas estadísticas como “Google Insights for Search” que permiten clasificar y evaluar las tendencias en distintas áreas de interés. Los trabajos precedentes se han centrado predominantemente en el mercado de trabajo, en los mercados de la vivienda en las ventas minoristas y en la confianza de los consumidores. En este trabajo se ha presentado una aplicación muy específica para la economía española, los flujos de turistas británicos entrados en España —los principales clientes de la industria turística española—. La mejora en la predicción que proporcionan los modelos de corto plazo que incluyen el G-indicador depende del modelo que se tome como referencia. Permite, no obstante, obtener un indicador ajustado del flujo de turistas británicos con casi un mes de adelanto.

Dada la importancia del turismo para la economía española, entre los próximos pasos a dar está la realización de estimaciones similares para turistas de otros países, de tal forma que se pueda mejorar la calidad predictiva de este tipo de modelos, así como obtener un indicador adelantado de los viajeros entrados y en última estancia del gasto por turismo. Otras aplicaciones a explorar son las ventas de automóviles, la confianza de los consumidores y la compra de viviendas. La principal característica de estos procedimientos es que, con el tiempo y los continuos avances en el uso de Internet, los resultados no pueden más que mejorar en el futuro. No obstante es necesario recordar que la construcción de estos G-indicadores ha de ser cuidadosa con el objeto de evitar equívocos derivados, entre otros aspectos, del distinto uso del lenguaje en diferentes países. Ignorar estas cautelas y confiar de forma ciega en estos indicadores que no siempre mejoran la capacidad predictiva de los modelos de previsión a corto plazo más tradicionales puede llevar a la obtención de resultados erróneos.

BIBLIOGRAFÍA

- ASKITAS, N. y K. ZIMMERMANN (2009). "Google Econometrics and Unemployment Forecasting", DIW Berlin.
- BAKER, S. y A. FRADKIN (2011). "What Drives Job Search? Evidence from Google Search Data", Department of Economics, 14 marzo, Stanford University.
- BOLLEN, J., H. MAOA y X. ZENGB (2010). "Twitter mood predicts the stock market". *Journal of Computational Science*, 2, pp. 1–8.
- BUGHIN, J. (2011). "Nowcasting the Belgian Economy". McKinsey.
- CAPORELLO, G. y A. MARAVALL (2004). Program TSW: Revised Reference Manual, Occasional Paper 0408, Research Department, Banco de España.
- CHAMBERLIN, G. (2010). "Googling the present", *Economic & Labour Market Review*.
- CHOI, H. y H. VARIAN (2009). "Predicting the Present with Google Trends", Google Research Blog.
- (2009). "Predicting Initial Claims for Unemployment Benefits", Google Research Blog.
- D'AMURI, F. y J. MARCUCCI (2009). "Google it! Forecasting the US unemployment Rate with a Google Job Search index", *Fondazione Eni Enrico Mattei*.
- DELLA PENNA, N. y H. HUANG (2009). "Constructing consumer Sentiment Index for U.S. Using Internet Search Patterns". Department of Economics, WP 2009-26, University of Alberta.
- EUROPEAN COMMISSION (2011). Survey on the attitudes of Europeans towards tourism, Wave 3.
- EUROSTAT (2011). ICT usage in households and by individuals.
- FEDERAL RESERVE OF NEW YORK (2011). Sentiment Analysis and Social Media Monitoring Solution RFP.
- FONDEUR, Y. y F. KARAME (2011). "Can Google Data Help Now or Fore-casting French Unemployment?", (Mimeo).
- KARAMÉ GINSBERG, J. et al (2008). "Detecting influenza epidemics using search engine query data", *Nature*.
- KHOLODILIN, K., M. PODSTAWSKI y B. SILVERSTOV (2010). "Do Google Searches Help in Nowcasting Private Consumption? A Real-Time Evidence for the US", DIW Berlin, abril.
- MCLAREN, N., y R. SHANBHOGUE (2011). "Using Internet search data as economic indicators", *Quarterly Bulletin* 2011Q2, Banco de Inglaterra.
- MCKINSEY GLOBAL INSTITUTE (2011). Big data: The next Frontier for Innovation, Competition and Productivity.
- O'CONNOR, B., R. BALASUBRAMANYAN, B. R. ROUTLEDGE Y N. SMITH (2010). "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series", *Fourth International AAAI Conference on Weblogs and Social Media*.
- OECD (2011). OECD Telecommunications and Internet Statistics.
- (2010). OECD Tourism Trends and Policies.
- PREIS, T., D. REITH y H. STANLEY (2010). "Complex dynamics of our economic life on different scales: insights from search engine query data", *Phil. Trans. R. Soc. A* 2010 368, 5707-5719 doi: 10.1098/rsta.2010.0284.
- WEBB, G. KENT (2009). "Internet Search Statistics as a Source of Business Intelligence Searches on Foreclosure as an Estimate of Actual Home Foreclosures", *Information Systems Journal*.
- WU, L. y E. BRYNJOLFSSON (2009). "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales", MIT Sloan School of Management.

PUBLICACIONES DEL BANCO DE ESPAÑA

DOCUMENTOS OCASIONALES

- 0501 JOSÉ RAMÓN MARTÍNEZ-RESANO: Size and heterogeneity matter. A microstructure-based analysis of regulation of secondary markets for government bonds.
- 0502 ALICIA GARCÍA-HERRERO, SERGIO GAVILÁ Y DANIEL SANTABÁRBARA: China's banking reform: an assessment of its evolution and possible impact.
- 0503 ANA BUISÁN, DAVID LEARMONTH Y MARÍA SEBASTIÁ BARRIEL: An industry approach to understanding export performance: stylised facts and empirical estimation.
- 0504 ANA BUISÁN Y FERNANDO RESTOY: Cross-country macroeconomic heterogeneity in EMU.
- 0505 JOSÉ LUIS MALO DE MOLINA: Una larga fase de expansión de la economía española.
- 0506 VÍCTOR GARCÍA-VAQUERO Y JORGE MARTÍNEZ: Fiscalidad de la vivienda en España.
- 0507 JAIME CARUANA: Monetary policy, financial stability and asset prices.
- 0601 JUAN F. JIMENO, JUAN A. ROJAS Y SERGIO PUENTE: Modelling the impact of aging on Social Security expenditures.
- 0602 PABLO MARTÍN-ACEÑA: El Banco de Francia, el BPI y la creación del Servicio de Estudios del Banco de España a principios de la década de 1930. (Existe una versión en inglés con el mismo número).
- 0603 CRISTINA BARCELÓ: Imputation of the 2002 wave of the Spanish Survey of Household Finances (EFF).
- 0604 RAFAEL GÓMEZ Y PABLO HERNÁNDEZ DE COS: The importance of being mature: The effect of demographic maturation on global per-capita income.
- 0605 JUAN RUIZ Y JOSEP VILARRUBIA: Canales de reciclaje internacional de los petrodólares. (Existe una versión en inglés con el mismo número).
- 0606 ALICIA GARCÍA-HERRERO Y SERGIO GAVILÁ: Posible impacto de Basilea II en los países emergentes.
- 0607 ESTHER GORDO, JAVIER JAREÑO Y ALBERTO URTASUN: Radiografía del sector de servicios en España.
- 0608 JUAN AYUSO, ROBERTO BLANCO Y FERNANDO RESTOY: House prices and real interest rates in Spain.
- 0701 JOSÉ LUIS MALO DE MOLINA: Los principales rasgos y experiencias de la integración de la economía española en la UEM.
- 0702 ISABEL ARGIMÓN, FRANCISCO DE CASTRO Y ÁNGEL LUIS GÓMEZ: Una simulación de los efectos de la reforma del IRPF sobre la carga impositiva.
- 0703 YENER ALTUNBAS, ALPER KARA Y ADRIAN VAN RIXTEL: Corporate governance and corporate ownership: The investment behaviour of Japanese institutional investors.
- 0704 ARTURO MACÍAS Y ÁLVARO NASH: Efectos de valoración en la posición de inversión internacional de España.
- 0705 JUAN ÁNGEL GARCÍA Y ADRIAN VAN RIXTEL: Inflation-linked bonds from a central bank perspective.
- 0706 JAVIER JAREÑO: Las encuestas de opinión en el análisis coyuntural de la economía española.
- 0801 MARÍA J. NIETO Y GARRY J. SCHINASI: EU framework for safeguarding financial stability: towards an analytical benchmark for assessing its effectiveness.
- 0802 SILVIA IRANZO: Introducción al riesgo-país. (Existe una versión en inglés con el mismo número).
- 0803 OLYMPIA BOVER: The Spanish survey of household finances (EFF): Description and methods of the 2005 wave.
- 0804 JAVIER DÍAZ-CASSOU, AITOR ERCE-DOMÍNGUEZ Y JUAN J. VÁZQUEZ-ZAMORA: Recent episodes of sovereign debt restructurings. A case-study approach.
- 0805 JAVIER DÍAZ-CASSOU, AITOR ERCE-DOMÍNGUEZ Y JUAN J. VÁZQUEZ-ZAMORA: The role of the IMF in recent sovereign debt restructurings: Implications for the policy of lending into arrears.
- 0806 MIGUEL DE LAS CASAS Y XAVIER SERRA: Simplification of IMF lending. Why not just one flexible credit facility?
- 0807 MIGUEL GARCÍA-POSADA Y JOSEP M.ª VILARRUBIA: Mapa de exposición internacional de la economía española.
- 0808 SARAI CRIADO Y ADRIAN VAN RIXTEL: La financiación estructurada y las turbulencias financieras de 2007-2008: Introducción general. (Existe una versión en inglés con el mismo número).
- 0809 FRANCISCO DE CASTRO Y JOSÉ M. GONZÁLEZ-MÍNGUEZ: La composición de las finanzas públicas y el crecimiento a largo plazo: Un enfoque macroeconómico.
- 0810 OLYMPIA BOVER: Dinámica de la renta y la riqueza de las familias españolas: resultados del panel de la Encuesta Financiera de las Familias (EFF) 2002-2005. (Existe una versión en inglés con el mismo número).
- 0901 ÁNGEL ESTRADA, JUAN F. JIMENO Y JOSÉ LUIS MALO DE MOLINA: La economía española en la UEM: Los diez primeros años. (Existe una versión en inglés con el mismo número).

- 0902 ÁNGEL ESTRADA Y PABLO HERNÁNDEZ DE COS: El precio del petróleo y su efecto sobre el producto potencial. (Existe una versión en inglés con el mismo número).
- 0903 PALOMA LÓPEZ-GARCÍA, SERGIO PUENTE Y ÁNGEL LUIS GÓMEZ: Employment generation by small firms in Spain.
- 0904 LUIS J. ÁLVAREZ, SAMUEL HURTADO, ISABEL SÁNCHEZ Y CARLOS THOMAS: The impact of oil price changes on Spanish and euro area consumer price inflation.
- 0905 CORAL GARCÍA, ESTHER GORDO, JAIME MARTÍNEZ-MARTÍN Y PATROCINIO TELLO: Una actualización de las funciones de exportación e importación de la economía española.
- 1001 L. J. ÁLVAREZ, G. BULLIGAN, A. CABRERO, L. FERRARA Y H. STAHL: Housing cycles in the major euro area countries.
- 1002 SONSOLES GALLEGO, SÁNDOR GARDÓ, REINER MARTIN, LUIS MOLINA Y JOSÉ MARÍA SERENA: The impact of the global economic and financial crisis on Central Eastern and SouthEastern Europe (CESEE) and Latin America.
- 1101 LUIS ORGAZ, LUIS MOLINA Y CARMEN CARRASCO: El creciente peso de las economías emergentes en la economía y gobernanza mundiales. Los países BRIC.
- 1102 KLAUS SCHMIDT-HEBBEL: Los bancos centrales en América Latina: cambios, logros y desafíos.
- 1103 OLYMPIA BOVER: The Spanish Survey of Household Finances (EFF): description and methods of the 2008 wave.
- 1104 PABLO HERNÁNDEZ DE COS, MARIO IZQUIERDO Y ALBERTO URTASUN: Una estimación del crecimiento potencial de la economía española. (Existe una versión en inglés con el mismo número).
- 1105 ENRIQUE ALBEROLA, CARLOS TRUCHARTE AND JUAN LUIS VEGA: Central banks and macroprudential policy. Some reflections from the Spanish experience.
- 1106 SAMUEL HURTADO, ELENA FERNÁNDEZ, EVA ORTEGA Y ALBERTO URTASUN: Nueva actualización del modelo trimestral del Banco de España.
- 1107 PABLO HERNÁNDEZ DE COS Y ENRIQUE MORAL-BENITO: Eficiencia y regulación en el gasto sanitario en los países de la OCDE. (Existe una versión en inglés con el mismo número).
- 1201 ELOÍSA ORTEGA Y JUAN PEÑALOSA: Claves de la crisis económica española y retos para crecer en la UEM. (Existe una versión en inglés con el mismo número).
- 1202 MARÍA J. NIETO: What role, if any, can market discipline play in supporting macroprudential policy?
- 1203 CONCHA ARTOLA Y ENRIQUE GALÁN: Las huellas del futuro están en la web: construcción de indicadores adelantados a partir de las búsquedas en Internet.

BANCO DE ESPAÑA
Eurosistema

Unidad de Servicios Auxiliares
Alcalá, 522 - 28027 Madrid
Teléfono +34 91 338 6363. Fax +34 91 338 6488
E-mail: publicaciones@bde.es
www.bde.es